DEA Environnement Marin : processus stochastiques

Avner Bar-Hen

Table des matières

1 Processus stochastique

1.1 Définition

Un **processus stochastique** $\{X_t : t \in \mathcal{T}\}$ est une suite de variables aléatoires indexées par $\mathcal{T}$ à valeurs dans un ensemble $\mathcal{X}$.

$\mathcal{T}$ est **l'ensemble des indices**. Souvent t représente le temps mais t peut être de dimension multiple (par exemple la longitude et la latitude). Si $\mathcal{T}$ est à valeur discrète on parle de **processus à temps discret**. Si l'ensemble des valeurs de $\mathcal{T}$ est continu, on parle de **processus à temps continu**.

$\mathcal{X}$ est **l'ensemble des états du processus**. L'ensemble des états peut être continu ou discret mais nous nous limiterons au cas où ces états sont en nombre fini ou dénombrable.

En résumé, on peut dire que la caractéristique de base d'un processus stochastique est le fait que la loi de la variable X_t soit fonction de t.

La notion de processus élargit la notion de variable aléatoire. Une réalisation d'un processus est appelée **trajectoire**. C'est donc la suite des réalisations des variables aléatoires X_t.

Les réalisations d'une même variable aléatoire pouvant être différentes, les réalisations d'un même processus peuvent donner des trajectoires différentes.

1.2 *Exemples*

1. Le cours d'une action cotée en bourse au jour le jour :
   - X_t : valeurs (en francs) de l'action à la date t
   - $\mathcal{X}$: $\mathbb{R}^+$
   - $\mathcal{T}$: {jour de cotation}
2. Un exemple de processus stochastique dégénéré est l'échantillon i.i.d :
   - $\{X_t \ : \ t \geq 1\}$ $X_t \sim \mathcal{N}(0, 1)$
   - $\mathcal{X}$: $\mathbb{R}$
   - $\mathcal{T} = 1, \ldots, n$

   Un processus plus intéressant est le processus cumulant : $S_t = \sum_{i=1}^{t} X_i$. Nous reviendrons rapidement sur ce processus dans le prochain paragraphe.
3. Jeu de "Pile ou Face". Après chaque lancer, le joueur gagne 1F s'il obtient "Pile" et perd 1F s'il obtient "Face". La variable X_n représentant sa fortune après n tirages est un processus appelé **marche aléatoire** ou **processus de Bernouilli**. ($\mathcal{X} = \mathbb{Z}$ et $\mathcal{T} = \mathbb{N}$). On peut noter qu'en relation avec l'exemple 2, ce processus peut être vu comme un processus cumulant.
4. La température au sol à un instant donné dans une parcelle est un processus doublement indicé. La loi de X_{mn} dépend de la longitude et de la latitude : $\mathcal{X} = \mathbb{R}$ et $\mathcal{T} = \mathbb{R}^2$.
5. Le nombre de cellules dans une culture à la date t. On suppose que chaque cellule se divise en deux au bout d'une durée aléatoire de temps.

1.3 *Processus cumulant*

Le processus cumulant est un mécanisme essentiel dans beaucoup de processus et nous allons donc en présenter les bases.
Soit $\{X_n\}_{n \in \mathbb{N}} = \{X_0, X_1, X_2, \ldots\}$ une suite de variables aléatoires gaussiennes indépendantes centrées réduites, c'est-à dire $X_i \sim \mathcal{N}(0, 1)$. Soit :

$$S_n = \sum_{i=0}^{n} X_i$$

S_n est distribué selon une loi normale car c'est la somme de variables aléatoires gaussiennes indépendantes.

De plus :

$$E(S_n) = E\left(\sum_{i=0}^{n} X_i\right) = \sum_{i=0}^{n} E\left(X_i\right) = 0$$

$$V(S_n) = V\left(\sum_{i=0}^{n} X_i\right) = \sum_{i=0}^{n} V\left(X_i\right) = n$$

et donc

$$S_n \sim \mathcal{N}(0, n)$$

Il est important de noter que les S_n ne sont pas indépendants. En effet :

$$S_n = S_{n-1} + X_n$$

$$\begin{aligned}
\mathrm{Cov}(S_n, S_m) &= E(S_n S_m) - E(S_n)E(S_m) \\
&= E\left(\sum_{i=0}^{n}\sum_{j=0}^{m} X_i X_j\right) \\
&= \sum_{i=0}^{n}\sum_{j=0}^{m} E\left(X_i X_j\right) \\
&= \sum_{i=0}^{\min(n,m)} E\left(X_i^2\right) \\
&= \min(n+1, m+1)
\end{aligned}$$

On peut étendre cette définition à un processus à temps continu : on définit le processus $\{X_t : t \geq 0\}$ par

$$X_t \sim \mathcal{N}(0, 1)$$

$$\mathrm{Cov}(S_t, S_s) = \min(s, t)$$

Le processus ainsi défini est appelé **mouvement brownien**.
Historiquement ce processus est censé rendre compte de la trajectoire d'une particule dans un espace contenant d'autres particules.

2 Chaîne de Markov

2.1 Processus de Markov

Définition 0.1 *Un processus de Markov est un processus dont l'évolution future $\{X_s : s > t\}$ ne dépend de son passé qu'à travers son état à l'instant t :*

$$\forall s > t, \qquad \mathcal{L}(X_s|X_r : r \leq t) = \mathcal{L}(X_s|X_t)$$

où $\mathcal{L}(X_s|X_t)$ désigne la loi de X_s sachant X_t.
Cette définition signifie que, pour le futur, l'histoire du processus jusqu'à l'instant t est entièrement résumée par son état à l'instant t ; ou encore que le présent étant connu, le futur est indépendant du passé.

Revenons sur les exemples de la section précédente :
- le cours d'une action (exemple 1) n'est vraisemblablement pas un processus de Markov : la "mémoire" du processus est probablement plus longue (par exemple une tendance saisonnière) ;
- un processus dégénéré (exemple 2) est bien évidemment un processus de Markov, le processus cumulant aussi : seule compte la dernière valeur. On parle parfois d'**ordre** du processus ou encore de **mémoire**, c'est-à-dire de la "longueur" de la dépendance. Le processus cumulant est d'ordre 1 alors que le processus dégénéré est d'ordre 0.
- La fortune du joueur au jeu du "Pile ou Face" (exemple 3) est un processus de Markov si les tirages sont indépendants ;
- dans le cas de la température dans un champ (exemple 4) la question ne se pose pas par rapport à cette définition puisque l'indice est double.

2.2 Chaîne de Markov

Définition 0.2 *Une chaîne de Markov est un processus de Markov pour lequel $\mathcal{X}$ et $\mathcal{T}$ sont finis ou dénombrables.*

Une chaîne de Markov est donc un processus à temps discret.

En notant $i_1, i_2, \ldots$ les états contenus dans $\mathcal{X}$, nous avons :

$$\mathrm{P}(X_{n+1} = i_{n+1}|X_0 = i_0, \ldots, X_n = i_n) = \mathrm{P}(X_{n+1} = i_{n+1}|X_n = i_n)$$

La loi d'une chaîne de Markov est donc entièrement déterminée par les probabilités initiales $\pi_0(i) = \mathrm{P}(X_0 = i)$ et les **probabilités de transition** :

$$\pi_{n,n+1}(i, j) = \mathrm{P}(X_{n+1} = j|X_n = i)$$

Il est suffisant de donner les probabilités de transition en une étape car les autres probabilités s'en déduisent immédiatement :

$$\begin{aligned} \pi_{n,n+2}(i, j) &= \sum_{k \in \mathcal{X}} \pi_{n,n+1}(i, k)\pi_{n+1,n+2}(k, j) \\ \pi_{n,n+3}(i, j) &= \sum_{k \in \mathcal{X}} \pi_{n,n+2}(i, k)\pi_{n+2,n+3}(k, j) \\ &\text{etc.} \end{aligned}$$

On reconnaît là la formule de multiplication matricielle. Ce résultat se généralise :

$$\pi_{n,m+p}(i,j) = \sum_{k \in \mathcal{T}} \pi_{n,m}(i,k)\pi_{m,m+p}(k,j) \tag{1}$$

Cette équation est connue sous le nom d'**équation de Chapman-Kolmogorov**.

3 Chaînes de Markov homogènes

Définition 0.3 *On appelle chaîne de Markov homogène une chaîne de Markov dont les probabilités ne dépendent pas de l'instant n considéré :*

$$\pi_{n,n+1}(i,j) = \pi(i,j)$$

Par exemple la fortune du joueur au jeu de "Pile ou Face" est une chaîne de Markov homogène car la probabilité de passer d'une somme à une autre ne dépend pas de l'instant considéré (sous l'hypothèse de tirages indépendants...).
La loi d'une chaîne de Markov homogène est résumée dans la **matrice de transition** Π qui contient l'ensemble des probabilités de transition. Traditionnellement, l'indice de la ligne donne l'état au temps n et l'indice de la colonne donne l'état au temps $n+1$:

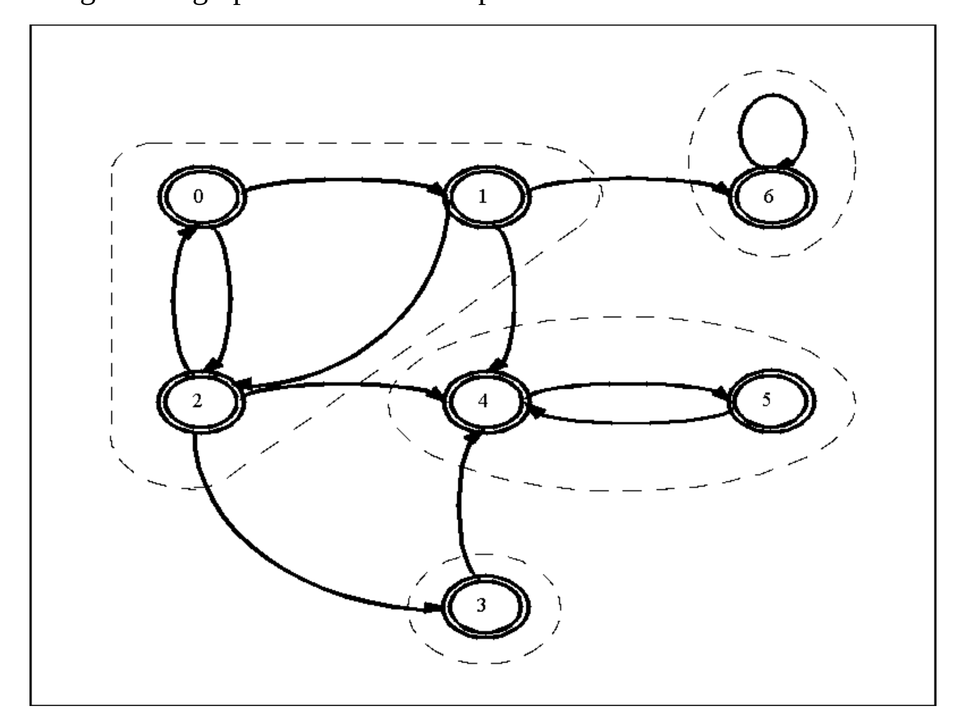
$$\Pi = \begin{pmatrix} \pi(1,1) & \dots & \pi(1,j) & \dots \\ \vdots & & \vdots & \\ \pi(i,1) & \dots & \pi(i,j) & \dots \\ \vdots & & \vdots & \end{pmatrix}$$

Tous les termes de la matrice des probabilités de transition Π sont positifs ou nuls et la somme des termes sur une ligne est égale à 1.En effet, quand on est dans un état donné, à l'étape suivante, on effectue une transition avec la probabilité 1. Les termes d'une ligne donnée constituent donc une loi de probabilité appelée **loi de transition** de l'état correspondant à l'indice de la ligne. Une matrice carrée possédant ces propriétés est appelée **matrice stochastique**.
Nous verrons plus loin (page 12) que la matrice de transition entre le temps n et le temps $n+k$ peut s'écrire Π^k.
Un cas particulier est celui des chaînes de Markov homogènes finies. La loi est donc donnée par une matrice de dimension finie.

3.1 *Classification des états*

Dans la définition que nous avons donnée d'une chaîne de Markov, l'évolution du processus au cours du temps à partir d'un état donné est entièrement décrite par la matrice des probabilités de transition. On peut aussi voir une chaîne de Markov comme un ensemble d'états entre lesquels s'effectuent des transitions. Certaines transitions sont possibles (probabilité de transition strictement positive) alors que d'autres sont impossibles

Figure 1 – graphe des transitions possibles d'une chaîne de Markov



(probabilité de transition nulle). Ceci nous amène à vouloir visualiser une chaîne de Markov en représentant chaque état par un sommet et chaque transition par un arc. Il faut noter qu'un arc possède une orientation. Ce point de vue structurel consiste en fait à visualiser le graphe des transitions possibles d'une chaîne de Markov (voir figure 1). Dans la mesure où les arcs sont orientés, on parle de **graphe orienté**. Si des poids sont associés aux arcs, on parle **d'automate**

Dans la théorie des graphes, on appelle **chemin** une succession d'arcs, telle que l'extrémité du $n^{\text{ème}}$ arc soit l'origine du $(n+1)^{\text{ème}}$ arc et on appelle **circuit** un chemin fermé. Le graphe des transitions possibles de la figure 1 comporte par exemple le chemin [0,1,2,3,4,5] et le circuit [0,1,2,0].

Pour la suite on note $\pi_n(i,j)$ la probabilité que le système soit dans l'état i au temps t et dans l'état j au temps $t+n$:

$$\pi_n(i,j) = \mathrm{P}(X_{t+n} = j | X_t = i) = \mathrm{P}(X_n = j | X_0 = i)$$

On dit que l'état j est **accessible** à partir de l'état i si la probabilité de passer de i à j est non nulle :

$$i \to j \Longleftrightarrow \exists n \geq 0 : \pi_n(i,j) > 0$$

En théorie des graphes ceci signifie qu'il existe un chemin entre i et j.

On dit que les états i et j **communiquent** si chacun d'eux est accessible à partir de l'autre :

$$i \leftrightarrow j \Longleftrightarrow \begin{cases} i \to j \\ j \to i \end{cases}$$

Pour que deux états ne communiquent pas il faut que l'un des deux ne soit pas accessible à partir de l'autre, c'est-à-dire :

$$\forall n \geq 0 \quad \pi_n(i,j) = 0 \text{ ou } \quad \forall n \geq 0 \pi_n(j,i) = 0$$

La relation de communication entre deux états est réflexive (par convention $\forall i \; \pi_0(i,i) = 1$), symétrique (par définition) et transitive, c'est donc une **relation d'équivalence**.
Il est donc possible de construire une partition des états d'une chaîne de Markov en classes d'équivalence telle que tous les états d'une classe communiquent entre eux et que deux états appartenant à deux classes différentes ne communiquent pas. Par construction, ces classes sont deux à deux disjointes et leur réunion est l'ensemble des états.
En théorie des graphes, une classe d'équivalence correspond à une **composante fortement connexe**, c'est-à-dire dont tous les éléments sont communiquants. On peut donc construire le **graphe réduit** (par exemple la figure 2). Dans ce graphe, les sommets représentent les classes et les arcs représentent les transitions possibles entre classes. Ce graphe possède la propriété d'être sans **circuit** (on ne peut jamais revenir au point d'origine), tous les circuits du graphe d'origine des transitions possibles ayant servi à construire les différentes classes.
Il est alors possible de distinguer deux types de classe :
- une classe est dite **transitoire** s'il est possible d'en sortir mais dans ce cas, le processus ne pourra plus jamais y revenir (classe (0,1,2) et classe (3) dans la figure 2) ;
- une classe est dite **récurrente** s'il est impossible de la quitter (classe (4,5) et classe (6) dans la figure 2).

Si une classe récurrente est composée d'un seul état, cet état est dit absorbant (état 6 dans la figure 2). Un état i absorbant est donc tel qu'une fois dans cet état on ne peut le quitter (par exemple la ruine dans le cas du jeu de "Pile ou Face"). En terme de probabilités de transition, ceci signifie que $\forall k \neq i$, $\pi_{ik} = 0$ et donc $\pi_{ii} = 1$.
Les états absorbants sont très particuliers puisqu'ils constituent des états terminaux du système. Il est notamment intéressant d'étudier les probabilités d'absorption, c'est-à-dire les probabilités que le système finisse par atteindre un tel état.
Les états d'une classe transitoire sont dits transitoires alors que les états d'une classe récurrente sont dits récurrents. Un état absorbant est donc un type particulier d'état récurrent.
Une chaîne de Markov pour laquelle il n'existe qu'une seule classe récurrente (égale à l'ensemble des états) est dite **irréductible**. Ceci signifie que tous les états communiquent.
Pour un état i de la chaîne, on appelle **temps de retour** le temps minimal pour revenir à l'état i ; c'est-à-dire le plus petit n tel que $\pi_n(i,i) > 0$.
Soit i un état d'une chaîne de Markov. La **période de retour** de i, notée T_i est la quantité définit par :

$$\begin{aligned} \text{Si } n = kT_i, k \in \mathbb{N} \quad &\Rightarrow \quad \pi_n(i,i) > 0 \\ \text{Si } n \neq kT_i, k \in \mathbb{N} \quad &\Rightarrow \quad \pi_n(i,i) = 0 \end{aligned}$$

c'est-à-dire que les retours à l'état i ne sont possibles que pour des durées multiples à la période.

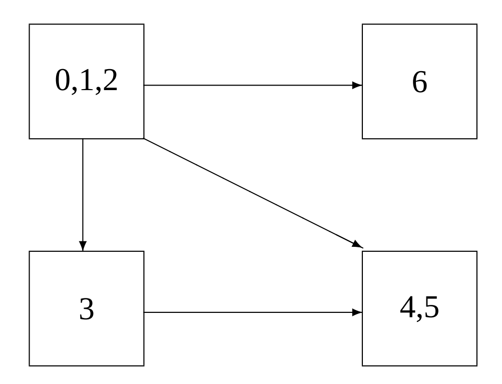


Figure 2 – exemple de graphe réduit

Une autre manière équivalente de dire les choses est de définir la période comme le pgcd$\{n \in \mathcal{T} : \pi_n(i,i) > 0\}$
L'état i est dit **périodique** si $T_i > 1$ et **apériodique** si $T_i = 1$.
Il est possible de montrer que deux états communiquants ont la même période et donc que la période est constante à l'intérieur des classes de communication.
La période commune des éléments de la classe est appelée **période de la classe**.
Si la chaîne est irréductible et qu'elle a une période, on parle d'une **chaîne périodique** ; si elle n'a pas de période on parle de **chaîne apériodique**.
Une chaîne irréductible et apériodique est dite **ergodique**.

Exemple 0.1 *Soit la matrice de transition de la chaîne :*

$$\Pi = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

la période de la chaîne est 2, le système agit comme un métronome.

3.2 Exemple : marche aléatoire

Un individu se déplace dans une direction fixe et peut, à chaque étape, soit faire un pas en avant (avec une probabilité p_i), soit faire un pas en arrière (probabilité q_i), soit rester sur place (probabilité $r_i = 1 - p_i - q_i$).
On suppose que ce processus est homogène, ce qui signifie que les probabilités des trois événements dépendent de l'endroit i où l'individu se trouve mais pas de l'étape n. En notant 0 le premier état, on obtient donc la matrice de transition :

$$\Pi = \begin{pmatrix} r_0 & p_0 & 0 & \dots & & \\ q_1 & r_1 & p_1 & 0 & \dots & \\ 0 & q_2 & r_2 & p_2 & 0 & \dots \\ \vdots & 0 & q_3 & r_3 & \ddots & \ddots \\ & \vdots & 0 & \ddots & \ddots & \\ & & \vdots & \ddots & & \end{pmatrix}$$

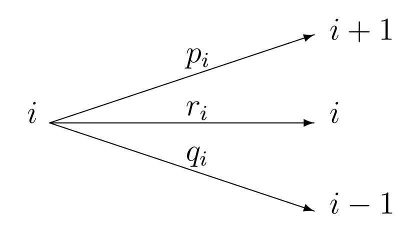


Figure 3 – Exemple de marche aléatoire

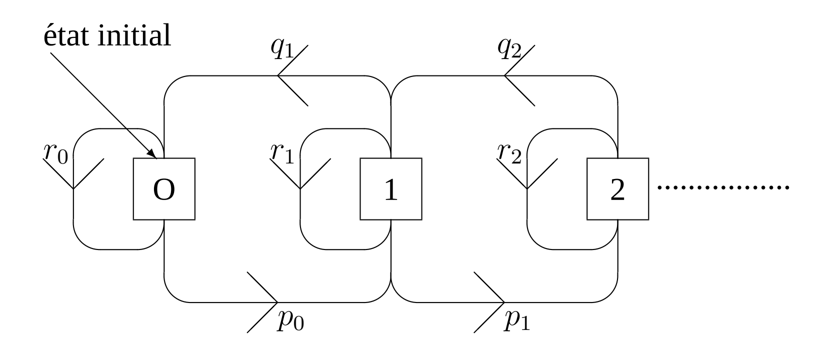


Figure 4 – Graphe d'une marche aléatoire

Cette matrice est de dimension finie ou infinie.
On peut représenter cette matrice sous forme d'un automate (voir figure 4)
De nombreux cas de marche aléatoire sont utilisés : fortune du joueur au jeu de "Pile ou Face", etc.

3.3 Comportement asymptotique

Dans la suite, on notera $\mu_n(i)$ la probabilité que le système soit dans l'état i à la $n^{\text{ème}}$ étape :

$$\mu_n(i) = \mathrm{P}(X_n = i)$$

On note $\mu_n = (\mu_n(1), \ldots, \mu_n(i), \ldots)$ avec $\sum_{i\in\mathcal{T}} = \mu_n(i) = 1$.
La définition des probabilités de transition implique :

$$\begin{aligned}\mu_n(i) &= \mathrm{P}(X_n = i) \\ &= \sum_{j\in\mathcal{T}} \mathrm{P}(X_n = i|X_{n-1} = j)\mathrm{P}(X_{n-1} = j)\end{aligned}$$

$$= \sum_{j \in \mathcal{T}} \pi(i,j)\mu_{n-1}(j)$$

Ce qui peut s'écrire matriciellement :

$$\mu_n = \mu_{n-1}\Pi$$

et donc de façon générale :

$$\mu_n = \mu_0\Pi^n$$

Cette équation correspond à l'écriture matricielle de l'équation 1.
Il est intéréssant de se demander si un processus donné finit par adopter un comportement stable ou pas, s'il converge vers une limite ou non. Pour cela on étudie le comportement asymptotique du vecteur μ_n.
Il est facile de montrer que la plus grande valeur propre de Π vaut 1.
On appelle **distribution stationnaire** la distribution de probabilité correspondant à tout vecteur propre de Π associé à la valeur propre 1 :

$$\mu = \mu\Pi$$

Le vecteur μ_n rend compte de la loi de X_n. Cette distribution est dite stationnaire si elle ne change pas lors d'une transition. Un tel vecteur μ rend compte d'un comportement stochastique stable du système.
Il est important de noter que μ n'est pas nécessairement unique.
On appelle **distribution limite** μ^* l'éventuelle limite de la suite μ_n.
Il est possible de montrer que si le vecteur μ_n admet une limite μ^*, alors μ^* correspond à une distribution stationnaire. Ceci implique que la limite μ^* ne dépend pas de l'état initial μ_0 du système.
Une condition nécessaire et suffisante pour l'existence d'une distribution limite indépendante de μ_0 est que 1 soit une valeur propre simple de Π (c'est-à-dire de multiplicité égale à 1) et que le module des autres valeurs propres soit strictement inférieur à 1. Ceci signifie qu'il n'existe qu'une seule classe récurrente. Une fois qu'on y rentre on ne peut plus en sortir.

3.4 Conclusion

Les sections précédentes n'ont permis que de soulever le coin du voile recouvrant la théorie des processus markoviens et il n'est pas dans notre but d'aller plus en avant sur le sujet. Les problèmes classiques de comptage consistent à estimer la loi du nombre d'occurrence d'un ou plusieurs états dans un temps donné. Nous renvoyons le lecteur intéressé à la bibliographie.

3.5 *Exemples d'application*

"Téléphone arabe"

Une histoire se transmet entre des individus par le phénomène du bouche à oreille. Il existe trois versions de cette histoire et chaque individu à qui est raconté une version a une probabilité p de la restituer telle quelle et une probabilité $q = 1 - p$ de la modifier en une des deux autres versions. La matrice Π s'écrit donc

$$\Pi = \begin{pmatrix} p & \frac{q}{2} & \frac{q}{2} \\ \frac{q}{2} & p & \frac{q}{2} \\ \frac{q}{2} & \frac{q}{2} & p \end{pmatrix}$$

et son polynôme caractéristique est

$$\begin{vmatrix} p-\lambda & \frac{q}{2} & \frac{q}{2} \\ \frac{q}{2} & p-\lambda & \frac{q}{2} \\ \frac{q}{2} & \frac{q}{2} & p-\lambda \end{vmatrix} = 0 \Rightarrow (1-\lambda)(p - \frac{q}{2} - \lambda)^2 = 0 \Rightarrow \begin{cases} \lambda_1 = 1 \\ \lambda_2 = \lambda_3 = p - \frac{q}{2} \end{cases}$$

On distingue deux cas :

1. $p - \frac{q}{2} = 1 \Rightarrow p = 1, \frac{q}{2} = 0$: la matrice Π est l'identité et le problème est trivial ;
2. $p - \frac{q}{2} < 1$: la seule valeur propre de module 1 est 1 et elle est simple. On se limite à ce cas.

La suite converge quel que soit μ_0 vers l'unique distribution stationnaire $\mu^* = (\mu_1^*, \mu_2^*, \mu_3^*)$ qui vérifie :

$$\mu = \Pi\mu \Longrightarrow \begin{cases} \mu_1^* = p\mu_1^* + \frac{q}{2}\mu_2^* + \frac{q}{2}\mu_3^* \\ \mu_2^* = \frac{q}{2}\mu_1^* + p\mu_2^* + \frac{q}{2}\mu_3^* \\ \mu_3^* = \frac{q}{2}\mu_1^* + \frac{q}{2}\mu_2^* + p\mu_3^* \end{cases} \Longrightarrow \mu^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$$

car $1 - p = q$

Ceci signifie qu'asymptotiquement il n'est pas possible de connaître la version originale. Pour bien comprendre la convergence du processus, il est intéressant de diagonaliser la matrice Π. Les vecteurs associés à $(p - \frac{q}{2})$ vérifient :

$$\left(p - \frac{q}{2}\right)\mu = \Pi\mu \Longrightarrow \begin{cases} (p - \frac{q}{2})\mu_1^* = p\mu_1^* + \frac{q}{2}\mu_2^* + \frac{q}{2}\mu_3^* \\ (p - \frac{q}{2})\mu_2^* = \frac{q}{2}\mu_1^* + p\mu_2^* + \frac{q}{2}\mu_3^* \\ (p - \frac{q}{2})\mu_3^* = \frac{q}{2}\mu_1^* + \frac{q}{2}\mu_2^* + p\mu_3^* \end{cases} \Longrightarrow \begin{cases} \mu^* = (1, -1, 0) \\ \mu^* = (0, -1, 1) \end{cases}$$

La diagonalisation de la matrice consiste à écrire Π sous la forme :

$$\Pi = P\Lambda P^{-1} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & p - \frac{q}{2} & 0 \\ 0 & 0 & p - \frac{q}{2} \end{pmatrix} \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ -1 & -1 & 2 \end{pmatrix}$$

et on a

$$\Pi^n = (P\Lambda P^{-1})^n = P\Lambda^n P^{-1}$$

or

$$\Lambda^n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \left(p-\frac{q}{2}\right)^n & 0 \\ 0 & 0 & \left(p-\frac{q}{2}\right)^n \end{pmatrix} \overset{n\to+\infty}{\longrightarrow} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

puisque $p-\frac{q}{2}<1$. Et donc

$$\lim_{n\to+\infty}\Pi^n = P\left(\lim_{n\to+\infty}\Lambda^n\right)P^{-1} = \frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Ce qui signifie que chaque personne à qui est racontée une version de l'histoire en restitue une autre tirée équiprobablement parmi les 3 versions.
Ce résultat est assez logique si l'on note que les états de la matrice Π sont non-identifiables, c'est-à-dire qu'on peut les intervertir.

Application à la génétique

La génétique est un des champs d'application privilégiés des chaînes de Markov car cela revient à supposer que l'information apportée par le passé du patrimoine génétique d'un individu est entièrement contenue dans le patrimoine génétique de ses parents. Cette hypothèse est en général raisonnable.
S.Wright a étudié la fluctation de la fréquence des gènes. Considérons une population de taille N d'individus haploïdes. On suppose la taille constante au cours des générations. Le nombre total de gènes est de $2N$: j gènes seront de type a et $2N-j$ gènes seront de type A.
Soit $\{X_n : n \geq 0\}$ le nombre de gènes a à la $n^{\text{ème}}$ génération. $\{X_n : n \geq 0\}$ est une chaîne de Markov. L'espace des états contient $2N+1$ valeurs $\{0,1,2,3,\ldots,2N\}$.
Si on néglige la mutation et la sélection, un modèle simple pour calculer les probabilités de transition d'une génération à l'autre consiste à supposer que, si l'on a j gènes de type a à une génération donnée, chaque gène de la génération suivante est le résultat d'une expérience de Bernouilli de paramètre $p_j = \frac{j}{2N}$ et la probabilité d'obtenir k gènes de type a est donnée par :

$$\pi_{jk} = \mathrm{P}(X_{n+1}=k|X_n=j) = C_{2N}^k p_j^k (1-p_j)^{2N-k} \tag{2}$$

Dans ce modèle simple, il est important de remarquer que :

$$\pi_{j,j} = 1 \text{ pour } j=0 \text{ et } j=2N$$

Il ne peut y avoir de distribution limite car il existe deux états absorbants.
D'un point de vue génétique, ceci signifie que l'endogamie produit une sélection en faveur des homozygotes.
Il serait pertinent de connaître la vitesse de convergence vers les états absorbants.
Introduction de la mutation Pour améliorer le modèle on peut introduire la mutation sous forme de deux termes : α le taux de mutation de a en A et β le taux de mutation de

A en a. La probabilité de transition π_{jk} garde la même forme mais le p_j de l'équation 2 devient :

$$p_j = \frac{j}{2N}\alpha + \left(1 - \frac{j}{2N}\right)\beta$$

Si α et β sont strictement positifs alors 0 et $2N$ ne sont plus absorbants. Il est possible de montrer que l'on a alors une distribution limite.

Introduction de la sélection On peut également introduire la sélection dans le modèle de base en supposant que le gène a a un avantage sélectif sur le gène A représenté par un terme $(1+s)$. Dans ce cas, la probabilité de transition π_{jk} garde la même forme mais le p_j de l'équation 2 devient :

$$p_j = \frac{(1+s)j}{2N+js}$$

Si il y a j gènes de type a à la $n^{\text{ème}}$ génération, l'espérance du nombre de gènes a à la génération $(n+1)$ vaut :

$$\mathrm{E}(X_{n+1}|X_n = j) = 2Np_j = 2N\frac{(1+s)j}{2N+js}$$

et l'espérance du nombre de gène A à la génération $(n+1)$ vaut :

$$\mathrm{E}(X_{n+1}|X_n = j) = 2N(1-p_j) = 2N\frac{2N-j}{2N+js}$$

Le rapport des deux espérances vaut :

$$\frac{(1+s)j}{2N-j} = (1+s)\frac{\text{nombre de gènes } a \text{ à la génération } n}{\text{nombre de gènes } A \text{ à la génération } n}$$

ce qui rend compte de la pression de sélection en faveur du gène a.

4 Processus de Poisson

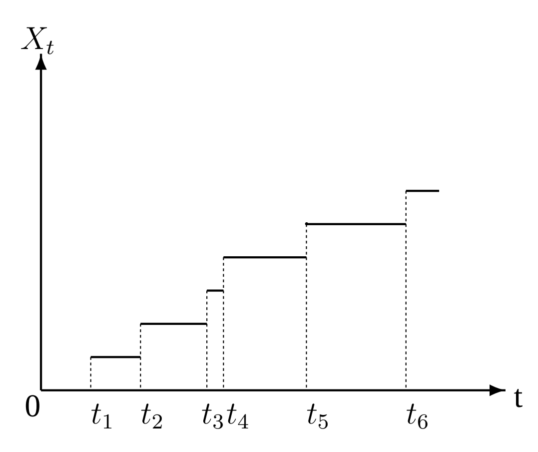
Dans la section 2, nous nous sommes intéressés aux concepts de base des chaînes de Markov et donc d'un processus à temps discret. Le but de cette section est de présenter (de manière rapide) un exemple important de processus à temps continu.

Un problème classique est de compter le nombre d'occurrence d'un événement donné dans un intervalle de temps. A titre d'exemple on peut citer les appels téléphoniques à un standard, l'occurrence d'accident à un carrefour ou l'apparition des bourgeons sur une plante.

La justification intuitive pour voir ces exemples comme des processus de Poisson provient de la loi des événements rares. Pour chaque petit intervalle de temps nous avons une expérience de Bernouilli dont la probabilité de succès est faible. Un résultat classique des statistiques permet de modéliser le nombre d'événements par une loi de Poisson. Nous reviendrons plus loin sur une justification plus formelle (mais que nous espérons tout aussi intuitive).

On note X_t le nombre d'événements survenus dans l'intervalle $]0, t]$.

La fonction de répartition est une fonction non-décroissante en escalier (voir figure 5).



$X(0) = 0$ X_t= nombre d'événements survenus avant t

Figure 5 – Trajectoire d'un processus de comptage

Définition 0.4 *X_t est est un* **processus poissonnien** *s'il vérifie les conditions suivantes :*

H1 *le processus est sans mémoire : les occurrences des événements sont indépendantes les unes des autres. Une autre manière équivalente de dire les choses est de postuler que l'occurrence d'événements avant la date t n'influe en rien sur l'occurrence d'événements après t ;*

H2 *le processus est homogène dans le temps : la loi de l'accroissement $(X_{t+h} - X_t)$ du processus ne dépend que de h et non pas de t (et est donc la même que celle de X_h).*

L'hypothèse (**H1**) induit que le processus de comptage des événements est un processus markovien : connaissant le présent, le futur est indépendant du passé.
Pour l'hypothèse (**H2**) on parle de **stationnarité** ou parfois d'homogénéité temporelle.
Par analogie avec la loi des événements rares, la probabilité d'observer plus d'un événement dans un intervalle de temps Δt tend vers 0 quand Δt tend vers 0. Cette propriété peut s'écrire :

$$\lim_{\Delta t \longrightarrow 0} \frac{\mathrm{P}(X_{t+\Delta t} - X_t > 1)}{\Delta t} = 0 \tag{3}$$

De manière équivalente, ceci peut s'écrire :

$$\mathrm{P}(X_{t+\Delta t} - X_t > 1) = o(\Delta t)$$

Divisons un intervalle de temps $]0, t]$ en N sous-intervalles de longueur $\Delta t = \frac{1}{N}$. La probabilité qu'un événement survienne dans un sous-intervalle vaut $1 - P(X_{\Delta t} = 0) - o(\Delta t)$ et donc l'espérance du nombre d'occurrence dans l'intervalle de temps vaut :

$$N(1 - P(X_{\Delta t} = 0) - o(\Delta t)) = \Delta t^{-1}(1 - P(X_{\Delta t} = 0) - o(\Delta t)) \tag{4}$$

Si Δt tend vers 0, et sous réserve de quelques conditions, l'équation 4 tend vers une limite correspondant au nombre d'occurrence de l'événements considéré dans un intervalle de temps de longueur t. Cette limite est appelée **l'intensité** et est notée λ :

$$\lim_{\Delta t \longrightarrow 0} \Delta t^{-1}(1 - P(X_{\Delta t} = 0)) = \lambda \tag{5}$$

Ce qui peut s'écrire :

$$P(X_{\Delta t} = 0) = 1 - \lambda \Delta t + o(\Delta t)$$

De même l'équation 3 peut s'écrire :

$$\lim_{\Delta t \longrightarrow 0} \Delta t^{-1}(1 - P(X_{\Delta t} = 0) - P(X_{\Delta t} = 1)) = 0$$

C'est-à-dire :

$$P(X_{\Delta t} = 1) = \lambda \Delta t + o(\Delta t)$$

4.1 *Système différentiel*

L'utilisation de système différentiel en processus est classique mais il est néanmoins possible de passer ce paragraphe lors d'une première lecture.
Pour Δt suffisamment petit, nous venons d'obtenir le système d'équations :

$$\begin{cases} \mathrm{P}(X_{t+\Delta t} - X_t > 1) &= o(\Delta t) \\ \mathrm{P}(X_{t+\Delta t} - X_t = 1) &= \lambda \Delta t + o(\Delta t) \\ \mathrm{P}(X_{t+\Delta t} - X_t = 0) &= 1 - \lambda \Delta t + o(\Delta t) \end{cases}$$

Notons

$$p_n(t) = \mathrm{P}(X_t = n)$$

Pour $n > 0$, on a

$$\begin{aligned} p_n(t + \Delta t) &= \mathrm{P}(X_t = n)\mathrm{P}(X_{t+\Delta t} - X_t = 0) \\ &\quad + \mathrm{P}(X_t = n - 1)\mathrm{P}(X_{t+\Delta t} - X_t = 1) + o(\Delta t) \\ &= p_n(t)(1 - \lambda \Delta t) + p_{n-1}(t)\lambda \Delta t + o(\Delta t) \\ &= p_n(t) + \lambda \Delta t(p_{n-1}(t) - p_n(t)) + o(\Delta t) \end{aligned}$$

D'où :

$$\frac{p_n(t + \Delta t) - p_n(t)}{\Delta t} = \lambda(p_{n-1}(t) - p_n(t)) + \frac{o(\Delta t)}{\Delta t}$$

et en passant à la limite, on obtient :

$$\begin{aligned} \lim_{\Delta t \longrightarrow 0} \frac{p_n(t + \Delta t) - p_n(t)}{\Delta t} &= \lambda(p_{n-1}(t) - p_n(t)) \\ p'_n(t) &= \lambda(p_{n-1}(t) - p_n(t)) \end{aligned}$$

Pour $n = 0$ on a :

$$\begin{aligned} p_0(t + \Delta t) &= \mathrm{P}(X_t = 0)\mathrm{P}(X_{t+\Delta t} - X_t = 0) \\ &= p_0(t)(1 - \lambda \Delta t + o(\Delta t)) \end{aligned}$$

D'où :

$$p_0'(t) = -\lambda p_0(t)$$

Les fonctions $p_n(t)$ vérifient donc le **système différentiel** :

$$\begin{cases} p_0'(t) &= -\lambda p_0(t) \\ p_n'(t) &= \lambda(p_{n-1}(t) - p_n(t)) \end{cases}$$

Il est possible de montrer que la solution de ce système est :

$$\begin{aligned} p_0(t) &= e^{-\lambda t} \\ p_1(t) &= \lambda t e^{-\lambda t} \end{aligned}$$

et par récurrence :

$$\forall n \geq 0 \quad p_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

4.2 *Interprétation des résultats*

La solution du système différentiel

$$\mathrm{P}(X_t = n) = p_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

signifie que, à tout instant t, la variable X_t suit une **loi de Poisson** de paramètre λt. On retrouve là une interprétation naturelle de l'intensité λ :

$$X_t \sim \mathcal{P}(\lambda t)$$

On en déduit immédiatement l'espérance et la variance de X_t :

$$\mathrm{E}(X_t) = \mathrm{V}(X_t) = \lambda t$$

Ce résultat peut aussi se retrouver en considérant **l'approche binomiale** déjà évoquée ; on découpe l'intervalle $[0, t]$ en N intervalles de taille $\frac{1}{N}$ suffisamment petit pour ne pouvoir contenir qu'au plus un seul événement et ce, avec une probabilité $\frac{\lambda t}{N}$. Dans chaque sous-intervalle la probabilité d'apparition d'un événement suit une loi de Bernouilli de paramètre $\frac{\lambda t}{N}$. Les intervalles étant indépendants, X_t correspond à la somme des N lois de Bernouilli et donc :

$$X_t \sim \mathcal{B}\left(N, \frac{\lambda t}{N}\right)$$

quand N tend vers l'infini, $\frac{\lambda t}{N}$ tend vers 0 et la loi binomiale tend vers la loi de Poisson et :

$$X_t \sim \mathcal{P}(\lambda t)$$

4.3 Comparaison avec un modèle déterministe

Si on voulait construire un modèle déterministe, on aurait la fonction :

$$x(t) = \text{nombre d'événements observés dans l'intervalle }]0, t]$$

l'équation différentielle correspondant aux hypothèses d'absence de mémoire et de stationnarité serait :

$$\frac{dx}{dt} = \lambda$$

avec la condition initiale $x(0) = 0$. On obtient donc l'équation :

$$x(t) = \lambda t$$

qui correspond au comportement "moyen" (c'est-à-dire en espérance) du processus de Poisson.

4.4 Temps d'attente

Loi de la durée entre deux événements

On s'intéresse maintenant à la durée (aléatoire) séparant deux occurrences de l'événement. On se place à une date t_0 et on va étudier la variable T :

$$T = \text{temps d'attente jusqu'à la prochaine occurrence}$$

On a :

$$\begin{aligned} \mathrm{P}(T > t) &= \mathrm{P}(X_{t_0+t} - X_{t_0} = 0) \\ &= \mathrm{P}(X_t = 0) \text{ hypothèse d'indépendance temporelle} \\ &= p_0(t) = e^{-\lambda t} \end{aligned}$$

La loi de T est donc indépendante de t_0 et on a

$$\mathrm{P}(T > t) = e^{-\lambda t} \Longleftrightarrow \mathrm{P}(T \leq t) = 1 - e^{-\lambda t}$$

c'est-à-dire T suit une loi exponentielle de paramètre λ :

$$T \sim \mathcal{E}(\lambda)$$

Il est important de remarquer qu'on ne se préoccupe pas de savoir si t_0 est une date d'occurrence. l'hypothèse d'indépence temporelle implique que la loi de T reste inchangée.

Interprétation et propriétés

X_t suit une loi de Poisson de paramètre λt. On a donc $\mathrm{E}(X_1) = \lambda$. Ceci signifie que λ représente le nombre moyen d'événements survenant dans une unité de temps. Nous retrouvons là le sens de l'intensité du processus de Poisson.
De même T suit une loi exponentielle de paramètre λ, on a donc $\mathrm{E}(T) = \frac{1}{\lambda}$. La durée moyenne séparant deux événements est donc égale à $\frac{1}{\lambda}$.
Il est possible de démontrer que la loi de Poisson est la loi du nombre d'événements survenant dans une unité de temps quand ces événements sont séparés par des durées exponentielles indépendantes.
On peut aussi s'intéresser à la **loi conditionnelle** :

$$\begin{aligned}
\mathrm{P}(T > s+t | T > s) &= \frac{\mathrm{P}(T > s+t, T > s)}{\mathrm{P}(T > s)} \\
&= \frac{\mathrm{P}(T > s+t)}{\mathrm{P}(T > s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\
&= e^{-\lambda t} \\
&= \mathrm{P}(T > t)
\end{aligned}$$

Ceci signifie que la loi exponentielle est “sans mémoire”. On peut montrer que cette propriété est caractéristique de la loi exponentielle.
Cette propriété est à l'origine du **paradoxe de l'autobus** (poissonnien) :
Si un usager attend un bus d'une ligne sur laquelle les passages suivent une loi de Poisson, la loi (et donc l'espérance) reste constante au cours du temps. Concrètement ceci signifie que si les bus passent en moyenne toutes les 30 minutes et que l'usager a déjà attendu 15 minutes, l'espérance du temps à attendre reste inchangée et est de 30 minutes.

$$\mathrm{E}(T) = \mathrm{E}(T - t > T > t) = \frac{1}{\lambda}$$

Date du n^{ème} événement

Soit T_n la date (aléatoire) à laquelle survient le n^{ème} événement :

$$T_1 \sim \mathcal{E}(\lambda)$$

et de façon générale :

$$T_n - T_{n-1} \sim \mathcal{E}(\lambda)$$

pour $n > 0$ $(T_0 = 0)$.
Donc T_n est la somme de n variables exponentielles de paramètre λ. Il est possible de montrer que la densité $f_n(t)$ de la variable aléatoire T_n s'écrit :

$$f_n(t) = \frac{\lambda^n}{(n-1)!} e^{-\lambda t} t^{n-1}$$

C'est-à-dire une loi gamma dont le paramètre n est entier. On parle aussi de **loi d'Erlang** et donc :

$$\begin{aligned} \mathrm{E}(T_n) &= \frac{n}{\lambda} \\ \mathrm{V}(T_n) &= \frac{n}{\lambda^2} \end{aligned}$$

5 Tour d'horizon de quelques processus

Nous venons de donner les bases de deux processus fondamentaux : les chaînes de Markov et les processus de Poisson. Il est important de noter que le mécanisme de construction est sensiblement différent entre les deux processus. Dans le cas des chaînes de Markov, on spécifie la dépendance entre les variables aléatoires alors que dans le cas des processus de Poisson, on travaille à partir de la loi inter-événement, c'est-à-dire l'intervalle de temps entre l'occurrence de deux événements, on parlera dans ce cas de processus de type **intervalle**. Une autre manière classique de construire un processus est le point de vue **comptage**, c'est-à-dire, en général, le nombre d'occurrence d'un état dans une séquence. Dans tout cette section il est important de garder en tête ces deux notions d'intervalle et de comptage.
Il n'est évidemment pas possible de donner une liste exhaustive de tous les processus mais il nous a semblé important de faire un tour d'horizon des principales approches. Nous commençons par quelques généralisations du processus de Poisson (processus de naissance, naissance et mort, branchement et file d'attente) pour arriver à la généralisation la plus globale (renouvellement) puis nous nous intéressons à des généralisations des chaînes de Markov (chaînes d'ordre r, semi-chaînes de Markov et chaînes de Markov cachées).
Le but est plus de présenter des problèmes que de les résoudre. Ceci peut paraître un peu frustrant mais connaître l'existence d'un outil est souvent aussi important que d'en maîtriser le mode d'emploi. Le lecteur intéressé peut se reporter à la bibliographie pour un traitement rigoureux et approfondi des processus présentés.

5.1 Processus de naissance

Dans un processus de Poisson, la probabilité d'un événement est indépendante du nombre d'événements qui se sont déjà produits. Cette hypothèse peut être irréaliste. Un exemple de ce phénomène est la reproduction des organismes vivants (d'où le nom du processus) où sous certaines conditions la probabilité d'une naissance est directement proportionnelle à la taille de la population à l'instant considéré. Un tel processus (qui est une généralisation des processus poissonnien) est aussi parfois appelé **processus de Yule**.
Prenons l'exemple de la division cellulaire.
Soit $N(t)$ le nombre de cellules dans une culture à la date t. On suppose que chaque cellule se divise en deux au bout d'une durée aléatoire distribuée exponentiellement. On suppose que les cellules ont toutes la même probabilité $\lambda\Delta t$ de se diviser durant un

intervalle de durée Δt. On reconnaît les hypothèses initiales du processus de Poisson. Une fois qu'une cellule est divisée, on considère qu'on a affaire à deux cellules nouvelles suceptibles de se diviser à leur tour.
On s'intéresse ici à la naissance de nouveaux individus et non à leur mort et l'on obtient donc une modélisation croissante de la taille de la population.
En suivant un raisonnement analogue à celui suivi pour le processus de Poisson on obtient les **équations de récurrence** :

$$\begin{aligned} \mathrm{P}(N(t+\Delta t) = N(t)) &= 1 - \lambda N(t)\Delta t + o(\Delta t) \\ \mathrm{P}(N(t+\Delta t) = N(t)+1) &= \lambda N(t)\Delta t + o(\Delta t) \\ \mathrm{P}(N(t+\Delta t) = N(t)+k, k>1) &= o(\Delta t) \end{aligned}$$

On retrouve des équations semblables à celles obtenues pour un processus poissonnien mais elles ne sont plus homogènes dans le temps puisque ces équations dépendent de l'effectif $N(t)$.
En notant

$$p_n(t) = \mathrm{P}(N(t) = n)$$

on peut obtenir une équation différentielle dont la solution s'écrit :

$$p_n(t) = \binom{n-1}{n-N_0} e^{-\lambda t N_0}(1-e^{-\lambda t})^{n-N_0}$$

on reconnait la loi binomiale négative de paramètre N_0 et $e^{-\lambda t}$:

$$N(t) - N_0 \sim \mathcal{NB}(N_0, e^{-\lambda t})$$

On en déduit :

$$\mathrm{E}(N(t)) = N_0 e^{\lambda t}$$

c'est-à-dire qu'avec ce modèle la croissance de la population est exponentielle en espérance.
De plus :

$$\mathrm{V}(N(t)) = N_0 e^{\lambda t}(e^{\lambda t} - 1)$$

d'où le coefficient de variation :

$$\mathrm{CV}(N(t))^{-1} = \frac{\mathrm{E}(N(t))}{\sqrt{\mathrm{V}(N(t))}} = \frac{N_0 e^{\lambda t}}{\sqrt{N_0 e^{\lambda t}(e^{\lambda t}-1)}} \overset{t\to\infty}{\longrightarrow} \frac{1}{\sqrt{N_0}}$$

ce qui signifie que la variabilité relative autour de l'espérance est faible sauf si la population initiale est particulièrement petite.
La croissance exponentielle de la population n'est évidemment pas toujours réaliste. Pour contrôler l'expansion on utilise souvent des modèles hétérogènes avec une intensité λ dépendant de l'effectif :

$$\lambda = \lambda(N)$$

On peut aussi introduire une limite supérieure M avec des *fonctions de freinage* de la forme :

$$\lambda(N) = \lambda\left(1 - \frac{N}{M}\right)$$

De manière analogue on peut construire un **processus de mort**. Et en suivant le même raisonnement on obtient :

$$N(t) - N_0 \sim \mathcal{NB}(N_0, e^{\mu t})$$

On peut remarquer que dans ces deux modèles, l'espérance est une fonction exponentielle du temps et donc qu'une transformation logarithmique la ramène à une fonction linéaire :

$$\begin{aligned}\text{processus de naissance :}\quad & \log(\mathbf{E}(N(t)) = \log(N_0) + \lambda t\\ \text{processus de mort :}\quad & \log(\mathbf{E}(N(t)) = \log(N_0) - \mu t\end{aligned}$$

On peut donc facilement estimer les paramètres N_0, λ, μ par une régression linéaire du logarithme de l'effectif sur le temps.
On note que le point de vue adopté est de type comptage. Il est cependant possible d'étudier des paramètres de type intervalle : temps avant la $n^{\text{ème}}$ naissance, temps entre deux naissance, etc.

5.2 *Processus de naissance et mort*

Une description réaliste du développement d'une population doit évidemment tenir compte à la fois des naissances et des morts des individus qui la composent. (Nous sommes encore dans une logique de type comptage).
Un modèle naturel consiste à combiner les deux modèles de la section 5.1.
En utilisant le même raisonnement et les mêmes notations, on obtient :

$$\begin{aligned}p_n(t+\Delta t) &= p_n(t) \times \text{P(aucune naissance ni mort durant } [t; t+\Delta t])\\ &\quad + p_{n-1}(t) \times \text{P(une naissance durant } [t; t+\Delta t])\\ &\quad + p_{n+1}(t) \times \text{P(une mort durant } [t; t+\Delta t])\\ &\quad + o(\Delta t)\end{aligned}$$

C'est-à-dire :

$$\begin{aligned}p_n(t+\Delta t) &= p_n(t) \times (1 - n(\lambda+\mu)\Delta t)\\ &\quad + p_{n-1}(t) \times (n-1)\lambda\Delta t\\ &\quad + p_{n+1}(t) \times (n+1)\mu\Delta t\\ &\quad + o(\Delta t)\end{aligned}$$

On en déduit l'équation différentielle :

$$p'_n(t) = -n(\lambda+\mu)p_n(t) + (n-1)\lambda p_{n-1}(t) + (n+1)\mu p_{n+1}(t) \tag{6}$$

mais sa résolution est particulièrement complexe dans le cas général, c'est-à-dire pour une taille initiale N_0 quelconque.
Pour $N_0 = 1$ (c'est-à-dire une population initiale de taille 1), la solution de l'équation différentielle 6 est :

$$\begin{aligned} p_0(t) &= \mu B(t) \\ p_n(t) &= (1 - \mu B(t))(1 - \lambda B(t))(\lambda B(t))^{n-1} \end{aligned}$$

avec

$$B(t) = \frac{1 - e^{(\lambda-\mu)t}}{\mu - \lambda e^{(\lambda-\mu)t}}$$

Conditionnellement au fait que la population n'est pas éteinte à la date t (probabilité $(1 - \mu B(t))$, $N(t)$ suit donc une distribution géométrique :

$$(N(t)|N(t) > 0) \sim \mathcal{G}(1 - \lambda B(t))$$

On utilise le résultat particulier pour $N_0 = 1$ pour étudier le comportement d'une population d'effectif initial N_0 quelconque en supposant que l'évolution de la population est équivalente à l'évolution de N_0 populations d'effectif initial 1.

5.3 Processus de branchement

Ces processus sont utilisés pour décrire l'évolution des populations, la croissance (ou la décroissance) de leur effectifs, les probabilités d'extinction, etc.
On considère qu'à la génération 0 on a 1 individu. Cet individu peut avoir des descendants qui constituent la génération 1. Chaque individu peut avoir des descendants qui constituent la génération 2.
On notera les liens entre les processus de branchement et les processus de naissance et mort.
Les exemples d'application de ces processus sont nombreux en physique, en épidémiologie, en généalogie ou encore en génétique.
La survivance des noms de famille est un des premiers exemples de ce processus. Sir Galton (fondateur de l'eugénisme et cousin de Darwin) posa le problème de l'extinction des noms de famille au cours des générations. Watson fut le premier à proposer une solution mathématique à ce problème. Dans ces modèles, les seuls descendants considérés sont les enfants mâles.
On note Y_n l'effectif à la $\text{n}^{\text{ème}}$ génération et $X_{j,n}$ le nombre de descendants du $\text{j}^{\text{ème}}$ individu à la $\text{n}^{\text{ème}}$ génération ($j = 1, \ldots, Y_n$).
Par définition, une génération est égale à la réunion des descendants de tous les individus de la génération précédente :

$$Y_n = X_{1,n-1} + \cdots + X_{Y_{n-1},n-1} = \sum_{j=1}^{Y_{n-1}} X_{j,n-1}$$

Il est donc possible de voir un processus de branchement comme la combinaison d'un processus markovien et d'un processus cumulant.

On suppose que les individus se reproduisent indépendamment les uns des autres et que le nombre de descendants suit une loi qui ne dépend ni de l'individu parent ni de la génération.

$$\{X_{n,j} : n \geq 0 \text{ et } 1 \leq j \leq Y_{n-1}\} \text{ i.i.d.}$$

Le phénomène suit une loi stable au cours du temps et indépendante des individus. On note

$$p_k = \mathrm{P}(X_{n,j} = k) = \mathrm{P}(X = k)$$

La loi de X est entièrement déterminée par les p_k (qui ne dépendent ni de n ni de j).
En général on peut supposer que le nombre de descendants suit une loi classique :

$$\begin{aligned} X \sim \mathcal{B}(m,p) &\implies p_k = C_m^k p^k (1-p)^{m-k} 0 \leq k \leq m \\ X \sim \mathcal{P}(\lambda) &\implies p_k = e^{-\lambda}\frac{\lambda^k}{k!} k \geq 0 \end{aligned}$$

Le choix de la loi résulte d'hypothèses sous-jacentes faites sur le mode de reproduction.
Classiquement on recherche la loi de Y_n, les probabilités d'extinction, le comportement asymptotique, etc. On peut aussi compliquer un peu le problème en supposant que les individus parents peuvent avoir des descendants de plusieurs types.

5.4 *Renouvellement*

La théorie du renouvellement a commencé avec l'étude des pannes et des remplacements de composants tels les ampoules électriques. Ensuite, il est apparu clairement qu'un nombre important de problèmes pouvait se ramener à ce formalisme.
Supposons que la durée de vie des ampoules soit une variable aléatoire. Une ampoule neuve est installée au temps initial. Soit X_1 la date d'occurrence de la première panne. On remplace immédiatement l'ampoule. La deuxième panne se produit en $X_1 + X_2$. De manière générale, la n^{ème} ampoule brûle au temps

$$S_n = \sum_{i=1}^{n} X_i$$

Si les variables aléatoires X_n sont indépendantes et identiquement distribuées, on parle de processus de renouvellement.
Deux points de vue analogues caractérisent le processus de renouvellement. On s'intéresse soit au processus S_n : temps avant la $n^{\text{ème}}$ panne, soit à $N(t)$: nombre de pannes dans l'intervalle de temps $]0,t]$.
Connaissant la loi de T, temps entre deux événements successifs, on cherche les propriétés de $N(t)$ et S_n.
On reconnait les liens entre le processus de renouvellement et le processus cumulant abordé à la section 1.3.
On peut par exemple s'intéresser à l'espérance de $N(t)$, c'est-à-dire le nombre attendu de renouvellements dans l'intervalle de temps $]0,t]$:

$$E(N(t)) = M(t)$$

Cette équation est appelée la **fonction de renouvellement**.

Soit T_i la durée (aléatoire) entre l'événement $i-1$ et l'événement i. Par exemple T_1 représente l'intervalle de temps entre l'instant 0 et le premier événement.
La loi du nombre d'événements se produisant dans l'intervalle $]0, t]$ se déduit des fonctions de répartition des lois des intervalles de temps entre l'instant 0 et le $n^{\text{ème}}$ événement par la relation suivante :

$$\begin{aligned} P(N(t) = n) &= P(N(t) \geq n) - P(N(t) \geq n+1) \\ &= P(T_1 + \cdots T_n \leq t) - P(T_1 + \cdots T_n + T_{n+1} \leq t) \end{aligned}$$

avec $n = \frac{t}{\sup(T)}, \ldots, \frac{t}{\inf(T)}$
où $\inf(T)$ représente l'intervalle minimum entre deux intervalles de temps et $\sup(T)$ l'intervalle maximum entre deux événements.
L'équation précédente signifie que si l'on a au moins n événements jusqu'à l'instant t alors l'instant où se produit le $n^{\text{ème}}$ événement est inférieur ou égal à t.
L'étude des processus de renouvellement dépasse notre propos mais il est possible de montrer qu'une majorité de processus correspond à des cas particuliers de processus de renouvellement. Par exemple, un processus de Poisson de paramètre λ est un processus de renouvellement dont la loi inter-événement suit une distribution exponentielle (loi gamma ou loi d'Erlang, voir page 20). Dans ce cas la loi de comptage suit une loi de Poisson. Un autre processus classique est le processus de Bernouilli (exemple 3 de la page 5). Il est possible de montrer que la loi inter-événement suit une loi binomiale négative et que la loi de comptage suit une loi binomiale.

5.5 *Files d'attente*

On s'intéresse ici au phénomène d'attente qui peut être ramené de façon générale au problème suivant : des clients se présentent dans un lieu donné pour obtenir un service.
Exemples :
- arrivée de clients à un distributeur automatique de billets (1 guichet, file d'attente infinie) ;
- arrivée de voitures à une station service (s pompes, files d'attente potentiellement infinie) ;
- appels téléphoniques à un standard (s lignes, pas de file d'attente) ;
- atterrissages d'avions sur un aéroport (s pistes, temps d'attente limité).

On peut voir un système de files d'attente comme la combinaison de deux processus de renouvellement (arrivée et service) relié par un mécanisme de passage (les files).

Notations

Pour décrire un tel système on adopte en général les **notations de Kendall** :

F **:** la loi du processus des arrivées dans le système ;

G **:** la loi de la durée des services rendu au(x) guichet(s) ;

s **:** le nombre de guichets ;

N **:** la capacité du système, c'est-à-dire le nombre de clients présents simultanément dans le système.

On obtient ainsi une notation générale permettant de classifier les différents phénomènes d'attente : un système d'attente est désigné par le quadruplet :

$$F/G/s/N$$

Si on suppose (et on le fait fréquemment) que les phénomènes étudiés vérifient la propriété de Markov (c'est-à-dire que l'information sur le futur est entièrement contenu dans le présent), les lois F et G sont notées M.

$$M/M/s/N$$

Quand la capacité du système est infinie ($N = \infty$) on omet le dernier terme et la notation de Kendall devient simplement :

$$F/G/s$$

la longueur maximale de la file est égale à la capacité du système moins le nombre de guichets :

$$n_{\max} = N - s$$

On utilise également deux autres notations classiques :
- λ : le taux d'arrivée (par unité de temps).
  $\frac{1}{\lambda}$ représente donc l'intervalle de temps moyen entre deux arrivées dans le système ;
- μ : le taux de service (par unité de temps).
  $\frac{1}{\mu}$ représente donc la durée moyenne d'un service.

Problèmes à résoudre

L'objet mathématique central dans l'étude d'un tel problème est le processus stochastique $\{X_t : t \geq 0\}$ où X_t représente le nombre d'individus présents dans le système à l'instant t.
On distingue deux régimes différents du système : le **régime transitoire** et le **régime permanent**. Dans de nombreux cas, après une phase initiale instable (ouverture des guichets), le système atteint une phase stable (milieu de journée).
Dans la phase initiale, la loi du nombre d'individus dans le système dépend du temps (régime transitoire) et elle se stabilise dans la phase stationnaire.
Le régime stationnaire n'existe pas toujours. On ne discutera pas ici des conditions d'existence de ce régime.
Dans le cas du régime transitoire, la loi du processus est caractérisé par les probabilités :

$$p_n(t) = \mathrm{P}(X_t = n)$$

En régime transitoire, ces probabilités dépendent évidemment des conditions initiales.

Les calculs en régime transitoire sont souvent très lourds et on se limite souvent à l'étude en régime stationnaire caractérisé par les probabilités limites

$$p_n = \lim_{t \to \infty} (p_n(t)) = p_n(\infty) = \mathrm{P}(X_\infty = n)$$

On ne s'arrêtera pas ici aux conditions d'existence de ces limites.
L'étude des processus de file d'attente permet de connaître des caractéristiques telles que :

- la durée moyenne d'attente ;
- la durée moyenne de séjour dans le système ;
- le nombre moyen d'individus présents dans le système ;
- le taux d'occupation des guichets ;
- le taux de clients non-servis.

On notera que les deux premières questions sont de type intervalle alors que les trois dernières sont de type comptage
Le système d'attente de base est le système **M/M/1**. On a vu que le processus stochastique vérifiant la propriété de Markov est le processus de Poisson. Dans le système M/M/1, on suppose donc que

- les arrivées se font selon un processus de Poisson $\mathcal{P}(\lambda)$;
- la durée de service suit une loi exponentielle $\mathcal{E}(\mu)$

Nous ne pousserons pas plus en avant l'étude des files d'attente.

5.6 *Processus ponctuel*

Ces processus sont particulièrement utiles pour étudier des problèmes de répartition spatiale. Ils seront traités dans le chapitre suivant.

5.7 *Généralisation d'une chaîne de Markov*

Chaîne de Markov d'ordre r

On dit qu'une chaîne de Markov est d'**ordre** r si l'état du processus à l'instant n ne dépend que des r états précédents :

$$\mathrm{P}(X_n = i_n | X_{n-1} = i_{n-1}, \ldots, X_1 = i_1) = \mathrm{P}(X_n = i_n | X_{n-r} = i_{n-r}, \ldots, X_{n-1} = i_{n-1})$$

Comme pour une chaîne de Markov d'ordre 1 (c'est-à-dire du type étudié dans la section 2), la chaîne est caractérisée par les probabilités initiales et les probabilités de transition.
Cette généralisation est plaisante mais il est important de noter que si la chaîne de Markov a J états possibles, il existe J^{r+1} probabilités de transition et donc le nombre de paramètres croît exponentiellement avec la mémoire du processus.

Semi-chaîne de Markov

Dans une semi-chaîne de Markov, on considère que la transition entre des états distincts correspond à une chaîne de Markov. Par contre la probabilité de rester dans un état n'est

plus déduite du modèle mais est explicitement spécifiée par une loi discrète d'occupation des états.
Pour une chaîne de Markov X_n la loi d'occupation de l'état j est donnée par

$$\begin{aligned} d_j(u) &= P(X_{n+u+1} \neq j, X_{n+u} = j, \ldots, X_{n+2} = j | X_{n+1} = j, X_n \neq j) \\ &= \pi(j,j)^{u-1}(1 - \pi(j,j)) \end{aligned}$$

la loi de l'occupation de l'état j, c'est-à-dire le temps où le système reste à l'état j, est donc une loi géométrique.
Une semi-chaîne de Markov à J états est donc définie par les paramètres suivants :

1. probabilité initiales : $\pi_0(j)$=P$(X_0 = j)$; $j = 1, \ldots, J$ avec $\sum_j \pi_0(j) = 1$
2. probabilités de transition : $\pi(i,j)$=P$(X_n = j | X_{n-1} = i)$ avec
   - $\sum_{j \neq i} \pi(i,j) = 1 \quad i,j = 1, \ldots, J$
   - $\pi(j,j) = 0 \quad \forall j = 1, \ldots, J$
3. loi d'occupation des états :

$$d_j(u) = \mathrm{P}(X_{n+u+1} \neq j, X_{n+u} = j, \ldots, X_{n+2} = j | X_{n+1} = j, X_n \neq j) \forall j = 1, \ldots, J$$

La dépendance n'est donc plus traduite explicitement dans la définition du modèle mais implicitement dans la définition des lois d'occupation des états. Les lois d'occupation des états sont par exemple des lois discrètes élémentaires (loi binomiale, Poisson ou négative binomiale).
On peut interpréter le mécanisme d'une semi-chaîne de Markov comme suit : à un instant n donné, on passe de l'état i à l'état j selon une loi de transition de l'état i puis on reste dans l'état j un temps u qui suit la loi d'occupation de l'état j. Enfin, on effectue une nouvelle transition conformément à la loi de transition de l'état j.
Un problème spécifique se pose pour traduire la notion d'état absorbant. Un état absorbant est un état dans lequel, une fois rentré, on reste infiniment longtemps. Cette notion ne peut donc pas se traduire dans une loi d'occupation d'un état.
Il existe une relation forte entre les semi-chaînes de Markov et la théorie du renouvellement. Une semi-chaîne de Markov à deux états peut s'interpréter comme la combinaison de deux processus de renouvellement. La loi d'occupation du premier état est une loi inter-événement traduisant un phénomène donné alors que la loi d'occupation du deuxième état est une loi inter-événement traduisant un autre phénomène. Dans le cadre de la théorie du renouvellement, ce type de processus est appelé processus de renouvellement alterné.

Processus de Markov cachés

Une autre extension possible des chaînes de Markov consiste à faire l'hypothèse qu'une séquence discrète n'est pas directement obtenue par une chaîne de Markov mais indirectement par des lois de probabilité attachées aux états de la chaîne de Markov. Si $\{X_n : n > 0\}$ est une chaîne de Markov d'ordre 1, $\{Y_n : n > 0\}$ est une chaîne de

Markov cachée d'ordre 1 si la relation de dépendance suivante est vérifiée :

$$\begin{aligned}
&\mathrm{P}(X_n = i_n, Y_n = j_n | X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, Y_1 = j_1, \ldots, Y_{n-1} = j_{n-1}) \\
&= \mathrm{P}(X_n = i_n, Y_n = j_n | X_{n-1} = i_{n-1}) \\
&= \mathrm{P}(X_n = i_n | X_{n-1} = i_{n-1}) \mathrm{P}(Y_n = j_n | X_{n-1} = i_{n-1})
\end{aligned}$$

Les variables aléatoires Y_n ne dépendent donc que de l'état correspondant X_n. Les dépendances structurant le modèle sont par conséquent uniquement représentées au niveau du processus sous-jacent X_n.
Une chaîne de Markov cachée à I états, homogène dans le temps est définie par les paramètres initiaux :

- probabilités initiales

$$\pi_0(i) = \mathrm{P}(X_0 = i) \quad i = 1, \ldots, I \text{ et } \sum_j \pi_0(i) = 1$$

- probabilité de transition

$$\pi_n(i, j) = \mathrm{P}(X_n = j | X_{n-1} = i)$$

  avec $i, j = 1, \ldots, I$ et $\sum_j \pi_0(j) = 1$

Ceci définit une chaîne de Markov d'ordre 1 homogène dans le temps. Les I états de cette chaîne de Markov sous-jacente sont observés à travers les lois de probabilité suivantes :

- probabilités d'observation $\nu_n(a, b)$ $=\mathrm{P}(Y_n = b | X_n = a)$ avec $a, b = 1, \ldots, I$ et $\sum_b \nu_n(b) = 1$

Dans un processus markovien caché, on suppose que seules les variables aléatoires Y_n sont observables. C'est donc sur ces variables aléatoires qu'il faut raisonner pour comparer des caractéristiques théoriques à des caractéristiques observées (même si il est toujours possible de calculer les lois du modèle sous-jacent).
Ces techniques sont très utilisées en reconnaissance de la parole.

Références

[1] Cox D.R. (1962) : *Renewal Theory*. Chapman and Hall. London.

[2] Cox D.R. et Miller H.D. (1965) : *The theory of stochastic processes*. Chapman and Hall. London.

[3] Feller W. (1968) : *An introduction to probability and its applications.* Vol. I 3rd Edition. Wiley. New-York.

[4] Feller W. (1971) : *An introduction to probability and its applications.* Vol. II 2nd Edition. Wiley. New-York.

[5] Gordon P. (1965) :*Théorie de chaînes de Markov finies et ses applications*. Dunod. Paris.

[6] Guttorp P. (1995) : *Stochastic modeling of scientific data*. Chapman and Hall. London.

[7] Karlin S. et Taylor H. M. (1975) :*A first course in stochastic processes*. 2nd Edition. Academic Press. London.

[8] Karlin S. et Taylor H. M. (1981) :*A second course in stochastic processes*. Academic Press. London.